



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Susan J. Morris, Gillian E. Scott, and Keith N. Leppard

Article Title: Adenovirus late phase infection is controlled by a novel L4 promoter

Year of publication: 2010

Link to published article:

<http://dx.doi.org/10.1128/JVI.00107-10>

Publisher statement: None

1 Adenovirus late phase infection is controlled by a novel L4 promoter

2 Susan J. Morris, Gillian E. Scott and Keith N. Leppard *

3
4 Department of Biological Sciences, University of Warwick, Coventry, CV4 7AL, U.K.

5
6 * Author for correspondence: Keith N. Leppard

7 Phone: +44 24 7652 3579

8 Fax: +44 24 7652 3701

9 E-mail: Keith.Leppard@warwick.ac.uk

10
11 Short running title: adenovirus gene expression cascade

12
13
14 Abstract word count: 166

15 Word count (text): 4396

16 Number of figures and tables: 6

17
18
19 Key words: adenovirus, L4 proteins, late gene expression, early-late switch

20 **Abstract**

21 During human adenovirus 5 infection, a temporal cascade of gene expression leads ultimately to
22 the production of large amounts of the proteins needed to construct progeny virions. However the
23 mechanism for activation of the major late gene that encodes these viral structural proteins has
24 not been well understood. We show here that two key positive regulators of the major late gene,
25 L4-22K and L4-33K, previously thought to be expressed under the control of the major late
26 promoter itself, are initially expressed from a novel promoter that is embedded within the major
27 late gene and dedicated to their expression. This L4 promoter is required for late gene expression
28 and is activated by a combination of viral protein activators produced during the infection,
29 including E1A, E4 Orf3 and the intermediate phase protein IVa2, and also by viral genome
30 replication. This new understanding re-draws the long-established view of how adenoviral gene
31 expression patterns are controlled and offers new ways to manipulate that gene expression
32 cascade for adenovirus vector applications.

Introduction

Although years of study have produced a detailed understanding of most molecular events during human adenovirus type 5 (Ad5) infection (4, 27), how the transition in viral gene expression from the early to the late phase is controlled has remained poorly defined. This control is crucial since it determines the activity of the genes that encode virion proteins, and hence the productivity of the infection. Residual activity from these genes is a confounding factor in the utility of E1-deleted Ad5 vectors for long-term gene delivery (51).

Initial expression of E1A from the linear Ad5 genome provides transcriptional activators that, with host proteins, turn on expression of the remaining early genes E1B, E2, E3, E4 (Fig. 1A). The major-late transcription unit (MLTU) is also weakly active at this time, but only the most 5' proximal L1 product is produced (1, 35, 41). Around the time of transition to the late phase of infection, when replication of the viral genome also begins, transcription of intermediate genes IX and IVa2 commences (13, 39, 48, 49), while major-late promoter (MLP) activity greatly increases and its scope expands to direct expression of a full set of around 15 MLTU products from regions L1-L5 via alternative splicing and polyadenylation (35, 41). This transition in MLTU activity reflects transcriptional and post-transcriptional changes, both of which require proteins encoded in the MLTU L4 region (Fig. 1B). L4-22K and L4-33K act post-transcriptionally to activate the production of the full set of MLTU mRNAs (16, 33, 44). At the same time, the MLP is further activated by IVa2 protein (30, 45) working with L4-22K and/or L4-33K (2, 33, 38).

The essential role of L4-22K and L4-33K in producing full late-phase expression from the Ad5 MLTU creates a paradox since, according to the current model of Ad5 gene expression, their expression is only achieved as a consequence of this activation process. Here we show that a

novel Ad5 promoter expresses L4-22K and L4-33K independently from the MLP, resolving this paradox, and that this promoter is activated by a combination of viral proteins and viral DNA replication.

Materials and Methods

Plasmids. pTG3602-Ad5wt (pWT) is a clone of the complete wild type Ad5 genome (11); pTG3602-L4-22K⁻ (pL4-22K⁻) was derived from pWT and contains a premature stop codon within the C-terminal unique portion of the L4-22K ORF (38). Linear genome was prepared from each of these plasmids by *PacI* digestion. Specific protein expression plasmids pCMV-IX (9), pMEPCMV-IVa2 (7), pCMV22KFLAG, pCMV33KFLAG and pCMV100KFLAG (33), pcDNA3.1Orf3 and pcDNA3.1Orf3 N82A (21) have been described previously. pE1A, provided by J. Logan, contains Ad5 1-5788 cloned between the *EcoRI* and *SalI* sites of the pBR322 derivative, pML2, and with a deletion of the Ad5 *SacI* fragment 1770-5644. pcDNA3.1Orf6 contains the Ad5 E4Orf6 sequence (34089-33182), obtained by PCR and cloned at the *EcoRI* site of pcDNA3.1.

L4 luciferase reporter plasmids were generated by amplifying various fragments in the region of Ad5 25887-26295 using primer pairs containing restriction recognition sites for *KpnI* (5' primer) and *NheI* (3' primer) and cloning using these sites into pGL3-Basic luciferase reporter plasmid (Promega). pcDNA3.1HisLacZ (Invitrogen) was used as a transfection control. pA-22/33KFLAG was generated by amplifying the relevant sequence as an *EcoRI* fragment from Ad5 strain 300 wild-type viral DNA. pA-22KFLAG was generated from pA-22/33KFLAG by exchanging the *HindIII/EcoRI* 3' fragment (Ad5 26328 to 26785 and C-Terminal FLAG tag) with the equivalent fragment from pCMV22KFLAG (33).

78 pLoxPGFP was obtained by modification of pBiEGFPacI (16), with insertion of a LoxP
 79 sequence between the promoter P_{Bi}-1 and the EGFP ORF and a zeocin resistance gene driven by
 80 the thymidine kinase promoter. L4 shuttle plasmids were constructed from pBiEGFPacI (16) by
 81 first replacing the P_{Bi}-1 promoter and EGFP with a LoxP sequence and adding a hygromycin
 82 resistance cassette from pTK-Hyg (Clontech), and then inserting one of several L4 cassettes
 83 downstream of LoxP. These were: (a) TPL-L4 cassettes, comprising the Ad5 tripartite leader
 84 (6049-6089, 7111-7182, 9644-9733) and 109bp intron sequence downstream of leader 3 (9734-
 85 9842), joined to L4 sequence from either 198bp upstream of 100K ORF or 177bp upstream of
 86 22/33K ORF to a C-terminal FLAG tag on L4-33K giving pShuttle100/22/33KFLAG and
 87 pShuttle22/33KFLAG respectively; (b) L4-only cassettes comprising the L4 components of the
 88 TPL-L4 cassettes, giving pShuttle26018-22KFLAG and pShuttle26018-22/33KFLAG; or (c) the
 89 L4-22/33K ORF alone from its AUG at 26195, giving pShuttle22/33KFLAG ORF.

90 **Production of L4P⁻ genome.** pBR322Δ*Hind*III was generated by digesting pBR322 with *Hind*III,
 91 end-filling with DNA polI (Klenow fragment) and re-ligating. The *Nde*I fragment from pWT was
 92 subcloned into pBR322Δ*Hind*III to generate pWT*Nde*I19548-31088. The core promoter (26018-
 93 26098) was deleted from this subclone by a two-step PCR protocol using primers that
 94 incorporated sequence from each side of the deletion (Ad26003-26017/26099-26119f and
 95 Ad26112-26099/26017-25997r) with 3' and 5' flanking primers (Ad26511-26487r and Ad25198-
 96 26027f respectively). The second stage PCR product (25198-26511) with the core promoter
 97 deleted was digested with *Asc*I and *Hind*III at sites within the Ad5 sequence and used to replace
 98 the equivalent wild-type sequence, generating pWT*Nde*I19548-31088ΔP. The *Nde*I fragment
 99 from pWT*Nde*I19548-31088ΔP was then recloned into pWT to generate pL4P⁻, from which
 100 linear genome was prepared as above.

Cell Culture. 293 and HeLa cells were maintained in Dulbecco's MEM (DMEM) supplemented with 10% newborn bovine serum. 293TETOFF cells (Clontech) and their derivatives were maintained in DMEM plus 10% fetal bovine serum on plates precoated with polylysine. Full details of the isolation of L4 cell lines will be presented elsewhere. Briefly, 293TETOFF cells, preselected for stable expression of GFP following pLoxPGFP transfection and Zeocin selection, were transfected with either pShuttle100/22/33KFLAG or pShuttle22/33KFLAG together with 500ng Cre recombinase expression plasmid (pCre; Invitrogen) and hygromycin-resistant lines isolated.

Transfection and Inhibitors. Transient transfections were carried out in 12 well cultures with cells plated at a density of 7×10^5 cells/well, using TransIT-LT1 (Cambridge Bioscience) at a ratio of 3 μ l / μ g DNA following the manufacturers' protocols. For inhibition of ATM/ATR signalling, cells were pre-treated with 3mM caffeine for 3h, washed twice with serum-free medium and then transfected as above. 5h post-transfection, medium was replaced with medium containing 3mM caffeine until harvest 48h post-transfection. For inhibition of viral DNA replication in genome-transfected cells, 5h post-transfection media were replaced with media containing 10mM hydroxyurea until harvested 48h post-transfection.

Luciferase Reporter Assays. Assays were performed as previously described (33). β -galactosidase activity was used to correct luciferase levels for differences in transfection efficiency. The effect of added E1A on β -galactosidase expression in HeLa cells was adjusted for using the ratio of mean activity between cells expressing and not expressing E1A. Data are shown either as fold-induction, where mean corrected luciferase expression from the specified basic or parental reporter plasmid is set as 1, or as % activity of the defined L4 promoter reporter

(25887-26125). Values shown are the mean of triplicate independent determinations within an experiment, and are representative of multiple independent experiments.

Antibodies. Proteins were detected using the following antibodies: anti-FLAG rabbit polyclonal serum (Sigma) at 1:10,000 for western blotting and at 1:1000 for immunofluorescence; AbJLB1 rabbit polyclonal serum to Ad5 late proteins at 1:10,000 (16); rabbit anti-L4-100K (W. C. Russell, University of St Andrews) at 1:10,000 for western blotting and 1:1000 for immunofluorescence; rabbit anti-L4-33K at 1:1000 (18); mouse anti-DNA binding protein MAb B6-8 at 1:10,000 (40); anti-E4-Orf3 (6A11) rat monoclonal antibody at 1:500 (34); rabbit anti-IVa2 at 1:10,000 (7); and anti- β -tubulin mouse mAb (Sigma) at 1:200. Secondary antibodies were goat-anti-mouse IgG-horseradish peroxidase (HRP) conjugate (Sigma) at 1:5000, goat-anti-rabbit IgG-HRP (Santa Cruz) at 1:100,000, goat-anti-rat IgG-HRP (Chemicon) at 1:100,000 and Alexafluor594 goat-anti-rabbit IgG (Invitrogen) at 1:500.

Protein Expression Detection. Transfections used 500ng pA-22KFLAG or pA-22/33KFLAG alone or co-transfected with either 1 μ g linear viral genome or 500ng various expression plasmids. All transfections were equalised for DNA content by the addition of either salmon sperm DNA or empty vector to account for the absence of genome or expression plasmid respectively. Transfected cells were harvested 48h post-transfection and FLAG-tagged proteins isolated using anti-FLAG (M2) agarose as described previously (21). Samples corresponding to 50% volume of immunoprecipitated proteins and 1-2% volume of cell lysates taken prior to immunoprecipitation were resolved through either 10% or 15% SDS-polyacrylamide gels as appropriate. Proteins were transferred to ECL nitrocellulose membrane (GE Healthcare) and western blot analysis carried out as described (29) using horseradish peroxidase-conjugated secondary antibodies and detection

via ECL-Advance (GE Healthcare). Immunofluorescence was performed as previously described (28).

Results

Expression of Ad5 L4 22/33K and 100K proteins is separately regulated

Early data showing that L4 mRNAs accumulated before those from L2, L3 and L5 during the onset of the late phase suggested that L4 expression might be regulated differently from other MLTU regions although transcription rate data did not suggest any additional promoter within the MLTU (25). L4 encodes several proteins (Fig. 1B). During attempts to produce stable cell lines expressing L4-100K, 22K and 33K proteins from a shortened MLTU under control of a tetracycline-regulated promoter (Fig. 2A), cells were repeatedly obtained that showed correctly inducible 100K expression but constitutive expression of 33K (Fig. 2B). Similar constructs designed to express just the 22K and 33K proteins also gave only constitutively expressing cell lines (Fig. 2C) while the shuttle plasmids (Fig. 2D) used to produce these cell lines produced readily detectable 33K in transient assays despite lacking any known promoter (Fig. 2E, F). Removal from these constructs of all Ad5 sequence upstream of the L4-33K reading frame abolished this expression (Fig. 2E). These results suggested that L4 22/33K could be expressed from a novel viral promoter, independent of the Ad5 MLP.

Ad5 L4 contains a promoter for 22K/33K expression

The promoterless shuttle plasmids, which nonetheless expressed L4-33K, retained the Ad5 DNA that encodes the tripartite leader (TPL) which is spliced onto all MLP-encoded mRNAs during Ad infection. However, this was not required for L4-22/33K expression as plasmids lacking this TPL-encoding sequence were equally capable of expressing these proteins (data not shown). Attention therefore focused on the 177 bp of L4 DNA upstream of the 22/33K start codon also

present in these constructs, which was hypothesised to contain a promoter for L4 22/33K expression. When this sequence plus 100 bp of the downstream 22/33K reading frame (Fig. 3A, construct B) was placed in front of a promoterless luciferase reporter gene, it caused a modest 4-fold increase in activity (Fig. 3B). However, extending the sequence further upstream to position 25887 increased activation to 40-fold (Fig. 3B, construct E). Similar analysis of a series of constructs with different lengths of Ad sequence showed an essential sequence for the L4 promoter was located between 26018 – 26098 bp. This sequence alone activated luciferase expression more than 10-fold (Fig. 3B, construct D) and its deletion from the full promoter completely destroyed activity (Fig. 3C, construct I). Including sequence upstream of this essential sequence to position 25887 greatly increased promoter activity, while sequences downstream to 26125 were repressive; however, the presence of the upstream sequences overcame this repressive effect (Fig. 3B). The full L4 promoter (L4P), which gave >80-fold increase in activity over background, was therefore designated as Ad5 position 25887 – 26125. The addition of further downstream sequence, to position 26296, gave lower activity than the full promoter, possibly because this sequence includes the L4 22/33K AUG, which would reduce translation from the luciferase start codon. Alternatively, this decrease in activity may be due to negative control elements within the extended region.

Previously described Ad5 promoters are activated by the E1A proteins that are produced in the earliest stages of infection. These E1A proteins are constitutively expressed in 293 cells, where L4P activity was first detected (Fig. 3B). The importance of E1A to L4P activity was therefore tested in HeLa cells, in which basal activity of the promoter was very low. Activity was increased 40-fold by E1A, whereas adding further E1A in 293 cells gave only a 6-fold enhancement of a

much higher basal level (Fig. 3D). Thus the L4P is responsive to E1A. Co-transfection of wt Ad5 genome also strongly activated the promoter (Fig. 3D); this activation is explored further below.

L4 promoter is functionally relevant and required for infection to enter late phase

To determine whether L4P in its natural context directed expression of L4 proteins, DNA from position 25887 to the 3' end of the 22K coding sequence (FLAG tagged for detection and with a polyA site added) was cloned into a plasmid with no other viral or eukaryotic sequences present (pA-22KFLAG). This plasmid alone expressed little or no 22K-FLAG in 293 cells, but when co-transfected with 22K-deficient Ad5 genome 22K-FLAG was readily detected (Fig. 4A). Co-transfected genome also activated expression of 33K from a similar construct capable of expressing both 22K (untagged) and 33K-FLAG (Fig. 4B). Ad5 superinfection also activated L4P in HeLa cells (data not shown). To test whether these amounts of 22K expressed from L4P were functionally significant, we made use of a complementation assay using Ad5 22K⁻ genome. Such genomes are substantially defective in the synthesis of all proteins encoded in the MLTU but can be complemented *in trans* with a 22K expression plasmid (33). Using this test, L4P directed expression of functional amounts of 22K-FLAG, either from a 22K-only construct (Fig. 4C) or from a 22/33K construct (Fig. 4D). L4P is therefore active in its natural context and the amount of 22K it produces is sufficient to promote the early – late transition in MLTU activity.

To show the importance of L4P to the life-cycle of the virus, the sequence 26018-26098 (essential for L4P activity; Fig. 3C) was deleted within full length wild-type genome plasmid to create L4P⁻ genome. This deletion also disrupts the essential L4-100K open reading frame, however late proteins (except hexon) are still expressed from MLTU plasmids in the absence of 100K (16) meaning that the importance of L4P for gene expression from the MLTU could be tested using this mutant genome. The L4P⁻ genome expressed amounts of DBP, a representative

early protein, equivalent to wild-type genome, but it essentially failed to express either MLTU-derived proteins or IVa2, reproducing the phenotype of L4-22K⁻ genome (Fig. 4E). Cotransfection of 22K expression plasmid with L4P⁻ allowed significant expression of IVa2 and MLTU proteins, except hexon which accumulated when 100K was additionally expressed *in trans*. Thus, L4P⁻ genome is functionally deficient in L4-22K. Although the promoter deletion is some distance upstream of the major splice acceptor site by which L4-22K and 33K are expressed from the MLP (position 26158), its use could have been affected by the deletion, so preventing 22K/33K expression. However this is not the case since L4-33K expression from L4P⁻ genome was clearly induced following complementation with 22K *in trans* (Fig. 4E). Thus, the gross defect in late gene expression from L4P⁻ genome is due to lack of L4 promoter activity, which is therefore required to initiate late phase gene expression. Surprisingly, co-transfection of L4P⁻ with 100K alone did cause some increase in certain late proteins, particularly penton base, although substantially less than achieved by 22K complementation. This may reflect low levels of transcription from the MLP in the absence of L4-22K, from which translation can be enhanced by L4-100K. Alternatively, because, uniquely among the plasmids used here, the DNA present in the complementing L4-100K plasmid covers the L4P⁻ deletion, some rescue of the mutation may be occurring by homologous recombination in this case only.

Defining activators of the L4 promoter

The strong activation of the L4 promoter by Ad5 genome in cells that already expressed E1A suggested that genome supplied activators in addition to E1A. Full late phase activation of the MLTU requires genome replication so, given the expected timing of activation of L4P at the early-late phase transition, we first tested whether or not replication of the added genome was required for its effect. Genome fragmented to prevent its replication by exhaustive digestion with

BglIII+HindIII (digest 1: 23 fragments) still significantly activated L4P (Fig. 5A), suggesting that genome replication is not essential for L4P transactivation. However, replication must contribute to the activation caused by intact genome since: (a) digest 1 has somewhat lower activity than intact genome; and (b) hydroxyurea (HU), which blocks DNA synthesis among other effects, essentially blocked stimulation of the promoter by genome (Fig. 5B, D).

We considered the possibility that some component of L4P activation by complete or fragmented genome could be due to expression of specific viral proteins. Ad5 genome digest 2 (HaeII; 77 fragments), which truncates or separates from their promoters all known open reading frames except protein IX, was a much less potent activator of the L4 promoter than genome digest 1, which could, in principle, express E1A, IX, IVa2 and E4 Orf1 from individual fragments (Fig. 5A). Since additional E1A contributed only modestly to 22K expression from L4P in 293 cells (Fig. 5A), while E4 Orf1 is expressed late in Ad5 infection (14) and Orf1 mutants express late proteins normally (8, 43), the involvement of these proteins in genome-mediated L4P activation in 293 cells was unlikely. In contrast, production of IX and IVa2 begins around the early-late transition making them plausible L4P activators. When tested, IVa2 but not IX activated L4P significantly (Fig. 6A, B), although it was less potent than intact genome.

Ad5 E4 Orf3 and E4 Orf6 were also tested for effects on L4 protein expression, initially because these proteins had been shown to affect MLTU splicing during Ad infection (36, 37) and it was conceivable that part of the induction of 22K by genome was by altering the splicing balance between 22K and 33K expression. However, both 22K (from pA-22KFLAG) and 33K (from pA-22/33KFLAG) were induced by exogenous E4-Orf3 (Fig. 6C) suggesting this was not the case. Moreover, Orf3 activated L4P while Orf6 did not (Fig. 6A, E). Thus Orf3, like IVa2, is an

activator of L4P. Furthermore, this activity was retained by mutant Orf3 N82A (Fig. 6D), a protein which is defective in other Orf3 activities.

The fact that highly fragmented genome (digest 2: 77 fragments) was still able to activate L4P to some extent (Fig. 5A), even though it was incapable of replicating or expressing any viral proteins other than potentially IX, which did not activate in trans, suggested that a general stress / DNA damage response, consequent upon the transfection of this fragmented DNA, might be affecting the promoter. During infection, replicating linear adenovirus genome also activates this response. In support of this idea, a non-specific fragmented DNA also activated L4P (Fig. 5C). Fragmentation of the DNA was clearly the critical factor in this activating response since the same DNA was used unfragmented as the transfection control in all these experiments and had no effect on the promoter. Cellular responses to double-strand DNA breaks are signalled via the kinase ATM, which is inhibited by caffeine. However, caffeine treatment did not inhibit L4P activation by digested genome (Fig. 5D). Thus the mechanism by which L4P is activated by fragmented DNA remains unclear.

Discussion

Our data show that L4-22K and 33K expression is driven by a previously undetected intermediate-phase promoter (L4P) embedded within the body of the Ad5 MLTU, and that this promoter triggers progression of infection into the late phase. The ability of Ad5 to produce L4-22/33K proteins independent of the MLP resolves the paradox created by these otherwise MLP-derived proteins being required for their own production. This discovery therefore fills a significant gap in our understanding of how the Ad5 infectious cycle progresses to virus production.

As might have been expected from knowledge of other Ad5 promoters, L4P activity was strongly upregulated by several virus-derived factors. In particular, L4P was strongly activated by the presence and replication of viral genome. This stimulation was partly attributable to expression of viral protein activators, including E1A, E4 Orf3 and IVa2, synthesis of the latter itself being dependent on the onset of DNA replication (22). The role of E1A proteins was not surprising since they activate several other Ad promoters (5, 23). Similarly, the role of IVa2 in L4P activation fits with its induction at the onset of replication and with its role as a transcription factor in activating the MLP (32, 45). Since we have recently shown that L4-22K increases the levels of IVa2, probably through protein:protein interaction and stabilization (33), there is thus a mutual enhancement relationship between these two proteins. This creates a molecular switch that can re-enforce itself to drive the infection into the late phase.

The involvement of IVa2 in L4P activation may explain, at least in part, the increased activation of L4P that is observed when co-transfected genome is able to replicate, since expression of IVa2 is itself activated by replication (22). IVa2 expression may also account for some of the difference in L4P transactivation activity between digested genome preparations 1 and 2. Where the IVa2 gene remained intact (digest 1), the genome fragment preparation was still able to express low levels of IVa2 (data not shown) whereas equivalent amounts of digest 2 were, unsurprisingly, unable to express any IVa2 protein. However, there remains a component of activation by highly fragmented genome that cannot be accounted for by protein expression and which is not specific to the presence of Ad5 sequences. The basis of this component of L4P transactivation remains to be determined.

The relevance to L4P activation of known Orf3 activities, which either disrupt DNA damage responses through mislocalization of key cellular proteins (42), disrupt promyelocytic leukaemia

303 (PML) nuclear bodies through direct interaction with PML protein (10, 15, 21) to overcome
304 antiviral responses (46, 47), or bind the transcription factor TIF1 α (52), can be discounted since
305 these activities are lost by the Orf3 N82A mutant, which we have shown retains full activity
306 towards L4P. Two further Orf3 activities, regulation of MLTU RNA splicing (36, 37) and relief
307 of the repression of cellular p53 activity that is imposed by other viral proteins during the early
308 phase of infection (24), have not been tested for the N82A mutant, however Orf3 activation of
309 L4P was not diminished by a significant reduction in p53 levels (data not shown), suggesting that
310 Orf3 does not act via this route. Thus the basis for the activation of L4P by E4 Orf3 remains to be
311 determined.

312 Differences were observed between the activity of L4P when incorporated into 293-based cell
313 lines within an inducible expression cassette and when used transiently to drive L4-22K/33K
314 expression. In the former case, expression of L4 proteins (via their FLAG tag) was detectable
315 without additional inducers, whereas the plasmid-based promoter required induction by one or
316 more of several factors in order for its protein product to be detected. There are two potential
317 reasons for this difference. First, the level of protein detected in cell lines reflects accumulation to
318 steady state over a considerable time whereas in a transient assay the level of protein observed is
319 more dependent on its rate of synthesis. Second, the basal activity of L4P appears to be sensitive
320 to the state of the cells since, in some transient assays, a low level of 22K was detected from L4P
321 in the absence of genome and we have shown L4P induction by fragmented DNA, a known cell
322 stressor. Because the cell lines were maintained in a cocktail of drugs to maintain the appropriate
323 selection regime, this may also have imposed a stress on the cells that served to activate L4P.
324 There were also differences in the requirement for specific activators to observe detectable L4P
325 activity between the natural context where L4P was driving expression of L4-22K and the

326 luciferase reporter context. In the natural context, activity was highly dependent on activators
327 provided by viral genome while considerable luciferase reporter activity was seen in the absence
328 of activation, although this was further increased by trans-activation. We believe this difference is
329 likely to be due to the presence of negative regulatory elements downstream of our mapped
330 promoter region, within the 22/33K coding sequence.

331 The DNA sequence within the mapped L4P contains potential binding sites for a large number of
332 transcription factors, including known mediators of E1A activation such as E2F and ATF that
333 may well be significant in the activity of L4P as well as numerous other factors, the significance
334 of which cannot be predicted. It is notable that there is no obvious TATA box in the promoter
335 region and in this respect L4P is similar to the IVa2 promoter, which depends for its activity on
336 an initiator element at the transcription start site (12). However, the functional significance of this
337 lack of TATA box is uncertain since TATA boxes, although once thought to be fairly ubiquitous
338 features of eukaryotic promoters, have been more recently shown to be present in only around
339 25% of genes in genome-wide surveys (50).

340 L4P overlaps with the E2 late promoter (E2-L) on the opposite genome strand. Like L4P, E2-L
341 becomes active during the intermediate phase of infection (3). However, the two promoters have
342 different sequence requirements, with sequence from 25910-26065 providing E2-L full activity
343 with correct temporal regulation within the Ad replication cycle (6) while the minimum L4
344 promoter was shown here to be 26018-26098. Moreover, the two promoters differ in their
345 response to E1A, L4P being activated by E1A while E2-L is insensitive (26) or repressed (20).
346 Thus, L4P and E2-L are distinct and independent in their regulation.

347 The discovery of L4P creates a new understanding how the full late phase of Ad5 infection is
348 initiated which is important in two ways. First, it creates a new opportunity to enhance the

properties of Ad5-based gene delivery vectors. E1-deleted vectors have been intensively investigated for many applications, but in situations where long-term persistence of the delivered gene is required, cellular immune responses to residual viral gene expression products are a significant confounding factor (51). Whilst these problems can be avoided by removing all viral genes from the vector (31), such vectors are difficult to produce in quantity. Second, the importance of L4P can be speculatively linked to the ability of Ad5 to persist for extended periods in lymphoid cells in its natural host (17, 19). This property requires that Ad5 can regulate its infectious cycle to limit or prevent commitment to the late phase. The discovery of the essential role of L4P therefore presents further avenues through which to explore this aspect of Ad5 biology.

Acknowledgements

We thank D. Farley for development of the promoterless shuttle plasmid system employed here, and the following colleagues for their generous gifts of reagents: P. Hearing (pTG3602- L4-22K⁻, anti-33K antibody); S. J. Flint (pTG3602), T. Dobner (anti-Orf3 antibody) and W. C. Russell (anti-100K antibody).

This work was supported by the Biotechnology and Biological Sciences Research Council (grant numbers BBS/B/02169 and BB/E014550/1).

References

1. Akusjarvi, G., and H. Persson. 1981. Controls of RNA splicing and termination in the major late adenovirus transcription unit. *Nature* **292**:420-426.
2. Ali, H., G. LeRoy, G. Bridge, and S. J. Flint. 2007. The adenovirus L4 33-kilodalton protein binds to intragenic sequences of the major late promoter required for late phase-specific stimulation of transcription. *J. Virol.* **81**:1327-38.

- 372 3. **Baker, C. C., and E. B. Ziff.** 1981. Promoters and heterogeneous 5' termini of the
373 messenger-RNAs of adenovirus serotype-2. *J. Mol. Biol.* **149**:189-221.
- 374 4. **Berk, A. J.** 2006. Adenoviridae: the viruses and their replication, p. Chapter 63. *In* D. M.
375 Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, and M. A. Martin (ed.), *Fields Virology*,
376 5th Edition ed, vol. 2. Lippincott Williams and Wilkins.
- 377 5. **Berk, A. J., F. Lee, T. Harrison, J. Williams, and P. A. Sharp.** 1979. Pre-early
378 adenovirus-5 gene-product regulates synthesis of early viral messenger-RNAs. *Cell*
379 **17**:935-944.
- 380 6. **Bhat, G., L. Sivaraman, S. Murthy, P. Domer, and B. Thimmappaya.** 1987. In vivo
381 identification of multiple promoter domains of adenovirus E1a-late promoter. *EMBO J.*
382 **6**:2045-2052.
- 383 7. **Brey, S.** 1999. Construction of IVa2-deficient adenovirus using complementing cell lines.
384 PhD Thesis, University of Warwick, U.K.
- 385 8. **Bridge, E., and G. Ketner.** 1989. Redundant control of adenovirus late gene expression
386 by early region 4. *J. Virol.* **63**:631-638.
- 387 9. **Caravokyri, C., and K. N. Leppard.** 1995. Constitutive episomal expression of
388 polypeptide IX (pIX) in a 293-based cell line complements the deficiency of pIX mutant
389 adenovirus type 5. *J. Virol.* **69**:6627-33.
- 390 10. **Carvalho, T., J. S. Seeler, K. Ohman, P. Jordan, U. Pettersson, G. Akusjarvi, M.**
391 **Carmo-Fonseca, and A. Dejean.** 1995. Targeting of adenovirus E1A and E4-ORF3
392 proteins to nuclear matrix-associated PML bodies. *J. Cell Biol.* **131**:45-56.
- 393 11. **Chartier, C., E. Degryse, M. Gantzer, A. Dieterle, A. Pavirani, and M. Mehtali.** 1996.
394 Efficient generation of recombinant adenovirus vectors by homologous recombination in
395 *Escherichia coli*. *J. Virol.* **70**:4805-10.

- 396 12. **Chen, H., and S. J. Flint.** 1992. Mutational analysis of the adenovirus 2 IVa2 initiator
397 and downstream elements. *J. Biol. Chem.* **267**:25457-65.
- 398 13. **Chow, L. T., and T. R. Broker.** 1978. The spliced structures of adenovirus 2 fiber
399 message and other late mRNAs. *Cell* **15**:497-510.
- 400 14. **Dix, I., and K. N. Leppard.** 1993. Regulated splicing of adenovirus type 5 E4 transcripts
401 and regulated cytoplasmic accumulation of E4 mRNA. *J. Virol.* **67**:3226-3231.
- 402 15. **Doucas, V., A. M. Ishov, A. Romo, H. Juguilon, M. D. Weitzman, R. M. Evans, and**
403 **G. G. Maul.** 1996. Adenovirus replication is coupled with the dynamic properties of the
404 PML nuclear structure. *Genes Dev.* **10**:196-207.
- 405 16. **Farley, D. C., J. L. Brown, and K. N. Leppard.** 2004. Activation of the early-late
406 switch in adenovirus type 5 major late transcription unit expression by L4 gene products.
407 *J. Virol.* **78**:1782-1791.
- 408 17. **Fox, J. P., C. D. Brandt, F. E. Wassermann, C. E. Hall, I. Spigland, A. Kogon, and L.**
409 **R. Elveback.** 1969. The virus watch program: a continuing surveillance of viral
410 infections in metropolitan New York families. VI. Observations of adenovirus infections:
411 virus excretion patterns, antibody responses, efficiency of surveillance, patterns of
412 infectios, and relation to illness. *Am. J. Epidemiol.* **89**:25-50.
- 413 18. **Gambke, C., and W. Deppert.** 1981. Late nonstructural 100,000- and 33,000-dalton
414 proteins of adenovirus type 2. I. Subcellular localization during the course of infection. *J.*
415 *Virol.* **40**:585-593.
- 416 19. **Garnett, C. T., G. Talekar, J. A. Mahr, W. Huang, Y. Zhang, D. A. Ornelles, and L.**
417 **R. Gooding.** 2009. Latent species C adenoviruses in human tonsil tissues. *J. Virol.*
418 **83**:2417-2428.

- 419 20. **Guilfoyle, R. A., W. P. Osheroﬀ, and M. Rossini.** 1985. Two functions encoded by
420 adenovirus early region 1A are responsible for the activation and repression of the DNA-
421 binding protein gene. *EMBO J.* **4**:707-713.
- 422 21. **Hoppe, A., S. J. Beech, J. Dimmock, and K. N. Leppard.** 2006. Interaction of the
423 adenovirus type 5 E4 Orf3 protein with promyelocytic leukemia protein isoform II is
424 required for ND10 disruption. *J. Virol.* **80**:3042-3049.
- 425 22. **Iftode, C., and S. J. Flint.** 2004. Viral DNA synthesis-dependent titration of a cellular
426 repressor activates transcription of the human adenovirus type 2 IVa(2) gene. *Proc. Natl.*
427 *Acad. Sci. U. S. A.* **101**:17831-17836.
- 428 23. **Jones, N., and T. Shenk.** 1979. An adenovirus type 5 early gene function regulates
429 expression of other early viral genes. *Proc. Natl. Acad. Sci. U. S. A.* **76**:3665-9.
- 430 24. **Konig, C., J. Roth, and M. Döbelstein.** 1999. Adenovirus type 5 E4orf3 protein
431 relieves p53 inhibition by E1B-55-kilodalton protein. *J. Virol.* **73**:2253-2262.
- 432 25. **Larsson, S., C. Svensson, and G. Akusjärvi.** 1992. Control of adenovirus major late
433 gene-expression at multiple levels. *J. Mol. Biol.* **225**:287-298.
- 434 26. **Leff, T., and P. Chambon.** 1986. Sequence-specific activation of transcription by
435 adenovirus E1a Products is observed in HeLa-cells but not in 293 cells. *Mol. Cell. Biol.*
436 **6**:201-208.
- 437 27. **Leppard, K. N.** 2008. Adenoviruses: Molecular Biology, p. 17-23. *In* B. W. J. Mahy and
438 M. H. V. van Regenmortel (ed.), *Encyclopedia of Virology*, vol. 1. Elsevier, Oxford.
- 439 28. **Leppard, K. N., and R. D. Everett.** 1999. The adenovirus type 5 E1b 55K and E4 Orf3
440 proteins associate in infected cells and affect ND10 components. *J. Gen. Virol.* **80**:997-
441 1008.

29. **Lethbridge, K. J., G. E. Scott, and K. N. Leppard.** 2003. Nuclear matrix localization and SUMO-1 modification of adenovirus type 5 E1b 55K protein are controlled by E4 Orf6 protein. *J. Gen. Virol.* **84**:259-268.
30. **Lutz, P., and C. Keding.** 1996. Properties of the adenovirus IVa2 gene product, an effector of late phase-dependent activation of the major late promoter. *J. Virol.* **70**:1396-1405.
31. **McCormack, W. M., M. P. Seiler, T. K. Bertin, K. Ubhayakar, D. J. Palmer, P. Ng, T. C. Nichols, and B. Lee.** 2006. Helper-dependent adenoviral gene therapy mediates long-term correction of the clotting defect in the canine hemophilia A model. *Journal of Thrombosis and Haemostasis* **4**:1218-1225.
32. **Mondesert, G., C. Tribouley, and C. Keding.** 1992. Identification of a novel downstream binding protein implicated in late-phase-specific activation of the adenovirus major late promoter. *Nucleic Acids Res* **20**:3881-9.
33. **Morris, S. J., and K. N. Leppard.** 2009. Adenovirus serotype 5 L4-22K and L4-33K proteins have distinct functions in regulating late gene expression. *J. Virol.* **83**:3049-3058.
34. **Nevels, M., B. Tauber, E. Kremmer, T. Spruss, H. Wolf, and T. Dobner.** 1999. Transforming potential of the adenovirus type 5 E4orf3 protein. *J. Virol.* **73**:1591-1600.
35. **Nevins, J. R., and M. C. Wilson.** 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. *Nature* **290**:113-8.
36. **Nordqvist, K., K. Ohman, and G. Akusjarvi.** 1994. Human adenovirus encodes 2 proteins which have opposite effects on accumulation of alternatively spliced mRNAs. *Mol. Cell. Biol.* **14**:437-445.

37. **Ohman, K., K. Nordqvist, and G. Akusjarvi.** 1993. Two adenovirus proteins with redundant activities in virus growth facilitate tripartite leader mRNA accumulation. *Virology* **194**:50-58.
38. **Ostapchuk, P., M. E. Anderson, S. Chandrasekhar, and P. Hearing.** 2006. The L4 22-kilodalton protein plays a role in packaging of the adenovirus genome. *J. Virol.* **80**:6973-6981.
39. **Persson, H., U. Pettersson, and M. B. Mathews.** 1978. Synthesis of a structural adenovirus polypeptide in the absence of viral DNA replication. *Virology* **90**:67-79.
40. **Reich, N. C., P. Sarnow, E. Duprey, and A. J. Levine.** 1983. Monoclonal antibodies which recognize native and denatured forms of the adenovirus DNA-binding protein. *Virology* **128**:480-4.
41. **Shaw, A. R., and E. B. Ziff.** 1981. Transcripts from the adenovirus-2 major late promoter yield a single early family of 3' coterminal mRNAs and five late families. *Cell* **22**:905-916.
42. **Stracker, T. H., C. T. Carson, and M. D. Weitzman.** 2002. Adenovirus oncoproteins inactivate the Mre11-Rad50-NBS1 DNA repair complex. *Nature* **418**:348-352.
43. **Thomas, M. A., R. S. Broughton, F. D. Goodrum, and D. A. Ornelles.** 2009. E4orf1 limits the oncolytic potential of the E1B-55K deletion mutant adenovirus. *J. Virol.* **83**:2406-2416.
44. **Tormanen, H., E. Backstrom, A. Carlsson, and G. Akusjarvi.** 2006. L4-33K, an adenovirus-encoded alternative RNA splicing factor. *J. Biol. Chem.* **281**:36510-7.
45. **Tribouley, C., P. Lutz, A. Staub, and C. Kedinger.** 1994. The product of the adenovirus intermediate gene IVa2 is a transcriptional activator of the major late promoter. *J. Virol.* **68**:4450-4457.

- 488 46. **Ullman, A. J., and P. Hearing.** 2008. Cellular proteins PML and Daxx mediate an innate
489 antiviral defense antagonized by the adenovirus E4 ORF3 protein. *J. Virol.* **82**:7325-7335.
- 490 47. **Ullman, A. J., N. C. Reich, and P. Hearing.** 2007. Adenovirus E4 ORF3 protein inhibits
491 the interferon-mediated antiviral response. *J. Virol.* **81**:4744-4752.
- 492 48. **Wilson, M. C., N. W. Fraser, and J. E. J. Darnell.** 1979. Mapping of RNA initiation
493 sites by high doses of uv irradiation: evidence for three independent promoters within the
494 left 11% of the Ad-2 genome. *Virology* **94**:175-184.
- 495 49. **Winter, N., and J.-C. d'Halluin.** 1991. Regulation of the biosynthesis of subgroup C
496 adenovirus protein IVa2. *J. Virol.* **65**:5250-5259.
- 497 50. **Yang, C., E. Bolotin, T. Jiang, F. M. Sladek, and E. Martinez.** 2006. Prevalence of the
498 initiator over the TATA box in human and yeast genes and identification of DNA motifs
499 enriched in human TATA-less core promoters. *Gene* **389**:52-65.
- 500 51. **Yang, Y., F. A. Nunes, K. Berensci, E. E. Furth, E. Gonzol, and J. M. Wilson.** 1994.
501 Cellular immunity to viral antigens limits E1-deleted adenoviruses for gene therapy. *Proc.*
502 *Natl. Acad. Sci. U. S. A.* **91**:4407-4411.
- 503 52. **Yondola, M. A., and P. Hearing.** 2007. The adenovirus E4 ORF3 protein binds and
504 reorganizes the TRIM family member transcriptional intermediary factor 1 alpha. *J. Virol.*
505 **81**:4264-4271.

Figures

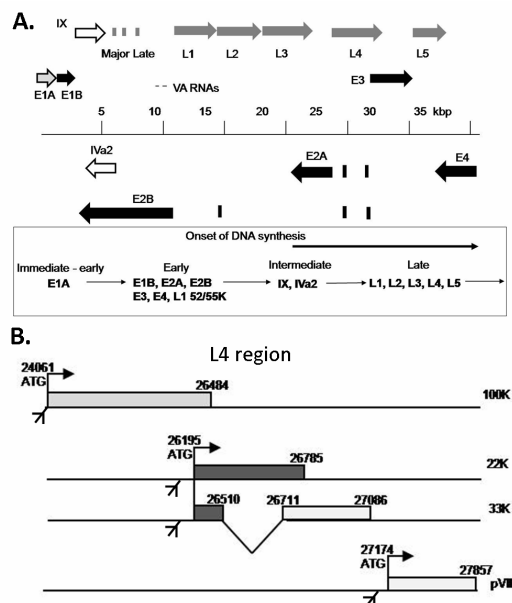


Figure 1. (A) The Ad5 transcription map showing immediate early (light grey), early (black), intermediate (white) and major late transcription units (MLTU; dark grey arrows), which are expressed during infection in a temporal cascade (bottom). All transcription units except IX and IVa2 produce multiple mRNAs and protein products by alternative RNA processing. MLTU mRNAs each comprise the three exons of the tripartite leader (TPL) spliced to one of ~15 possible acceptor sites within regions L1-L5. (B) Organisation of the L4 region of the MLTU showing L4-100K, L4-22K, L4-33K and pVIII ORFs, with 3' splice sites for their expression from the major late promoter. L4-22K and 33K share the same N-terminal sequence but have distinct C-termini.

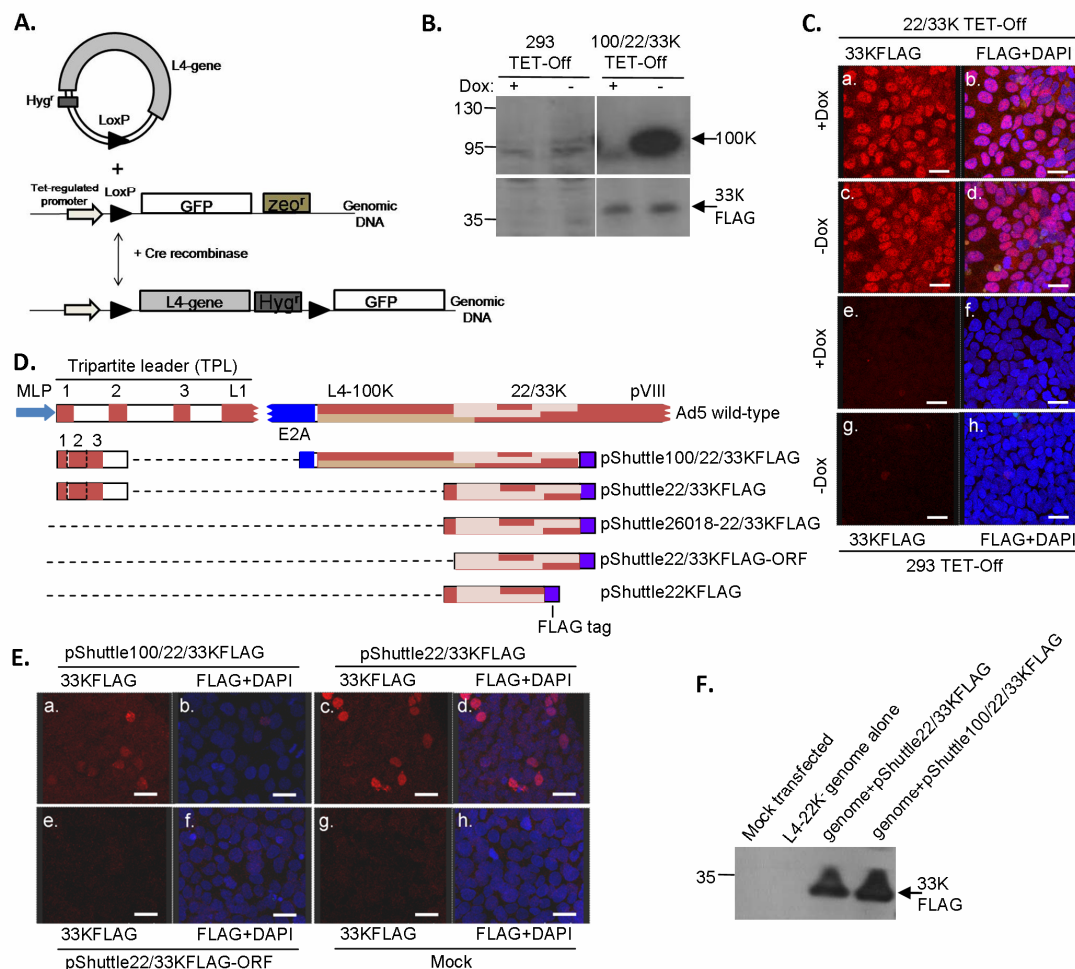


Figure 2. Differences in control of L4-22K, L4-33K and L4-100K expression. (A) Schematic representing the method for generating stable L4-protein cell lines. Target cells express GFP under doxycyclin control and this gene is then replaced with L4 sequences by Cre-mediated recombination with a promoterless shuttle plasmid. (B) 100/22/33KFLAG cells containing the L4 region linked to a tetracycline-regulated promoter or parental 293TETOFF cells were grown in the presence (+) or absence (-) of doxycycline for 3 days with daily media changes. L4-100K and L4-33KFLAG in total cell lysates were detected by western blot analysis. (C) 22/33KFLAG cells (a-d), or parental 293TETOFF cells (e-h) were grown in the presence (a, b, e, f) or absence (c, d, g, h) of doxycycline for 3 days with daily media changes. Cells were fixed and stained for

529 FLAG-tagged 33K (red) and for nuclear DNA (DAPI; blue) and images collected sequentially
530 using a Leica SP2 confocal microscope, to avoid cross-talk between the fluors. FLAG and DAPI
531 images were overlaid using Leica software (b, d, f, h); scale bar 20µm. (D) Schematic
532 representation of the Ad5 sequences present in pShuttle plasmids. Brown: MLTU exon sequences
533 (rightwards transcription); blue: E2A exon (leftwards transcription); pale brown: L4 ORFs as
534 indicated. (E) Transient expression from promoterless L4 shuttle plasmids. 293TETOFF cells
535 were transfected with pShuttle100/22/33KFLAG (a, b), pShuttle22/33KFLAG (c, d) or
536 pShuttle22/33KFLAG ORF (e, f), or mock transfected (g, h). Cells were fixed and stained 48h
537 later, and imaged as for panel C; scale bar 20µm. (F) 293TETOFF cells were mock-transfected or
538 transfected with L4-22K⁻ genome, together with either pShuttle100/22/33KFLAG or
539 pShuttle22/33KFLAG. 33KFLAG protein in total cell lysates was detected by western blot
540 analysis. Proteins molecular mass markers migrated to the positions shown on the left of panels B,
541 F (kDa).

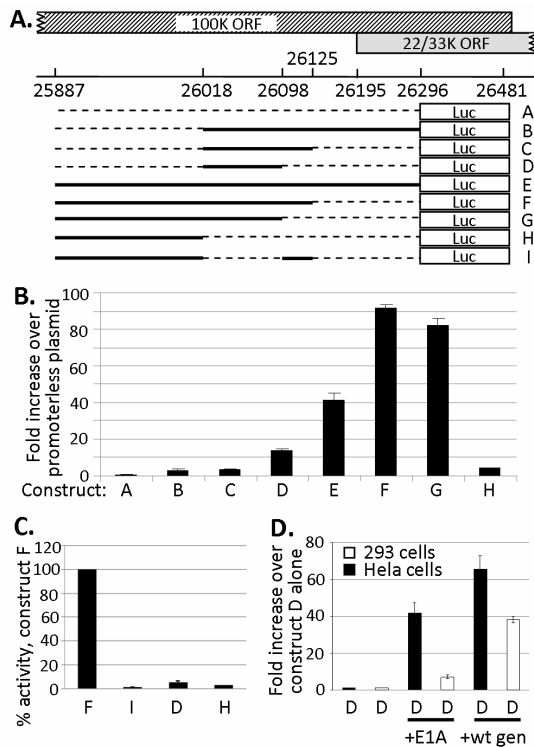


Figure 3. Defining sequences important for L4 promoter activity. (A) A schematic representation of the L4 sequences included in luciferase reporter constructs analysed in panels B-D. (B, C) 293 cells were mock-transfected or transfected with luciferase reporter constructs containing as promoter varying lengths of L4 sequence (panel A). Firefly luciferase activity, corrected for transfection efficiency using β -galactosidase expression from an independent control plasmid, is expressed (B) as fold difference from the activity of the promoterless reporter plasmid (pGL3Basic) or (C) as % of the activity of the full L4 promoter, construct F. (D) HeLa cells (black bars) or 293 cells (white bars) were transfected with construct D plus either empty vector, E1A expression plasmid or Ad5 wt genome (wt gen). Firefly luciferase activity, corrected as for (B, C), is expressed as fold difference from the activity of construct D in the presence of empty vector, which was set as 1. Each panel shows the mean values from biological triplicates within a single experiment (error bars, +/- standard deviation), and is representative of at least 3 independent experiments.

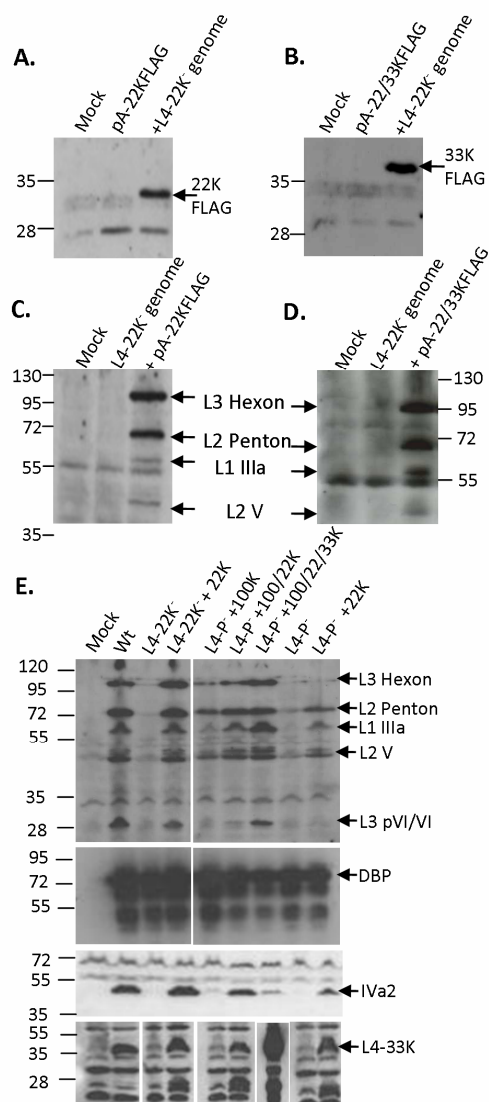


Figure 4. L4-22K expression from L4 promoter is functionally significant. (A - D) 293 cells were mock-transfected or transfected with L4-22K⁻ genome alone or plus either pA-22KFLAG (A, C) or pA-22/33KFLAG (B, D). Immunoprecipitated FLAG-tagged 22K (A) or 33K (B), or Ad5 late proteins in total cell lysates (C, D), were detected by western blot analysis. (E) Ad5 wild-type (Wt), L4-22K⁻ or L4P⁻ genomes were transfected into 293 cells with combinations of expression plasmids for individual L4 proteins 100K, 22K and 33K as indicated. Total cell extracts were analysed as for (A, B) for specific viral proteins indicated to the right of the panel.

564 Where a blot with a given antibody is shown in segments, these were all taken from the same
565 exposure of the same blot, with lanes rearranged for clarity of explanation. The positions to
566 which proteins of known molecular mass migrated are indicated for all panels (kDa).

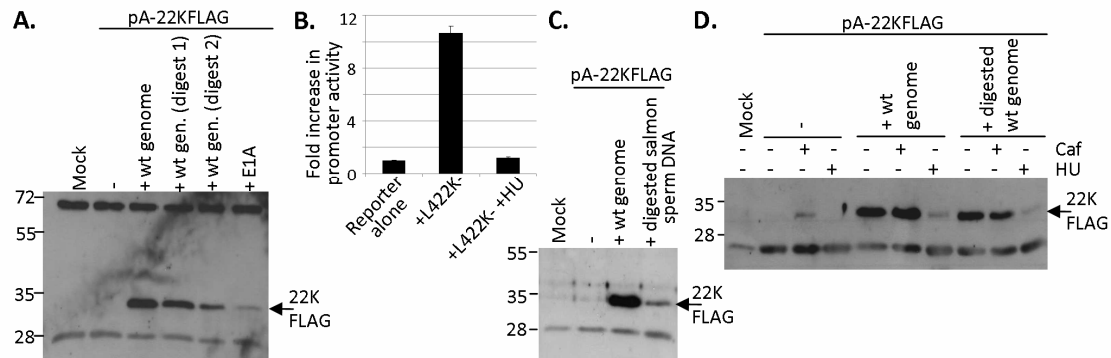


Figure 5. Activation of L4P by Ad5 genome. (A) 293 cells were mock transfected or transfected with pA-22KFLAG together with undigested Ad5wt genome (wt gen.), Ad5wt genome digested with either *Bgl*II/*Hind*III (digest 1) or *Hae*II (digest 2) or an E1A expression plasmid and 22KFLAG detected as for Figure 4. (B) 293 cells were transfected with L4 promoter luciferase reporter construct D (Fig. 3A), with or without L4-22K⁻ genome and/or treatment with 10mM hydroxyurea (HU) from 5h post-transfection. Luciferase expression, corrected for transfection efficiency, is expressed as fold increase over the activity of core promoter alone, set as 1. Error bars show the standard deviation of three replicate determinations. The graph is representative of two experiments. (C) As panel A except using digested salmon sperm DNA as an L4P activator in comparison with intact Ad5 genome. (D) 293 cells were mock treated or pre-treated with 3mM caffeine for 3h, then mock-transfected or transfected with pA25887-22KFLAG together with intact Ad5wt genome or digest 1. From 5h post-transfection cells were maintained in DMEM or DMEM supplemented with 3mM caffeine (Caf) or 10mM hydroxyurea (HU).

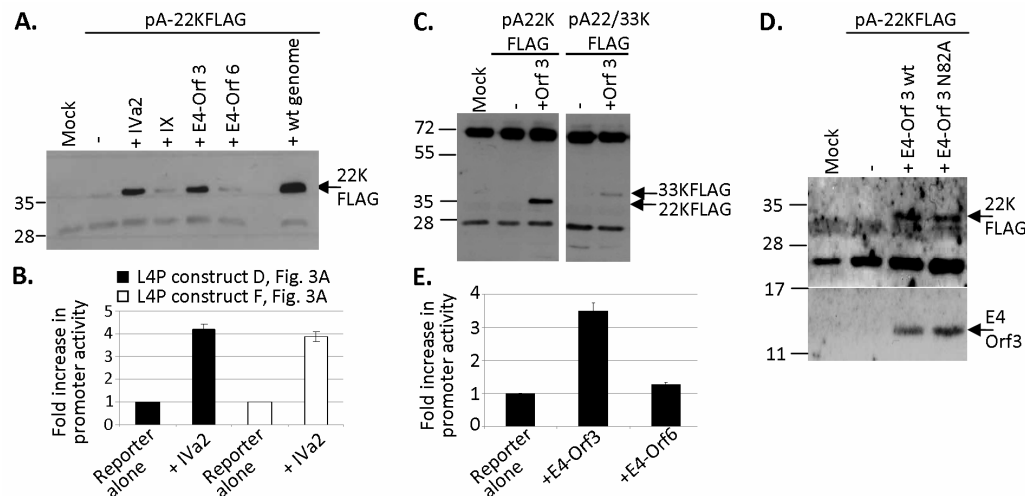


Figure 6. Viral proteins activate the L4 promoter. (A, C, D) 293 cells were mock transfected or transfected with pA-22KFLAG or pA-22/33KFLAG, together with viral protein expression plasmids as indicated or Ad5wt genome. Immunoprecipitated 22KFLAG protein or proteins from unfractionated cell lysates were detected by western blot analysis as indicated at the right of each panel. The positions of molecular mass marker proteins are shown on the left (kDa). (B) 293 cells were mock transfected or transfected with L4 promoter luciferase reporter constructs D or F (Fig. 3A), with or without IVa2 expression plasmid. Firefly luciferase activity, corrected for transfection efficiency, is expressed as fold difference from activity of the relevant promoter reporter alone, which was set as 1. Error bars show the standard deviation of three replicate determinations. (E) As panel B except cells were transfected with L4 promoter reporter construct D, with or without either E4-Orf3 or E4-Orf6 expression plasmids. The graphs are representative of two experiments.